# The Impact of different gamification types in the context of data literacy: An online experiment

Nikoletta-Zampeta Legaki[1], Daniel Fernández Galeote[1] and Juho Hamari[1]

[1] *Gamification Group, Faculty of Information Technology and Communication Sciences, Tampere University, Finland*

**Abstract**

As the pace at which humans create data increases, a new challenge for individuals and society is to turn a world full of data into a data-driven world. However, data and statistical literacy remain difficult topics to engage with whereas gamification rises as a promising technique to improve motivation. Therefore, we developed a software composed of interactive charts and tools aiming to teach data literacy in four different versions: (i) challenge- (badges), (ii) immersion- (avatars; story), and (iii) social-based (competition) gamification, along with (iv) a control version (no gamification) to compare the effects of different gamification types on learning outcomes. We conducted four-group random assignment pre-, post-test online experiments with students (N=181) from various courses, schools, and educational levels. The primary results of our experiments show a statistically significant improvement in students' performance of almost 44% from using the software. Gamification types did not result in statistically significant differences in students' learning outcomes, suggesting optimism regarding the contribution of interactive data visualization in improving data literacy.

**Keywords 1**

Gamification, data literacy, statistical literacy, education, exploratory data analysis

## 1. Introduction

We produce and consume data with great ease and frequency. Even a smartphone can process and visualize data easier than ever before. However, this data-explosion is as beneficial as the insights that we can get from it, and we still lack a fact-based view of our world [1]. Data literacy skills have become so critical that they have been suggested as a course in secondary education, highlighting the importance of "reading and writing with data" [2]. These skills help not only to understand the data around us, but also support a more rational approach to societal problems, realizing what is happening by using data, and eventually dealing rationally with societal challenges such as climate change or the COVID-19 pandemic. While we must address these challenges, we still lack motivation to gain data insights to transform a world full of data into a data-driven society [3].

Understanding data might lead to data-driven and well-informed decision-making at a personal or societal level [4]. Thus, there are online databases worldwide that provide data sets regarding a plethora of global topics (e.g., economy, the environment, etc.). Despite these initiatives, people are still discouraged from being statistically aware of this data, resulting in a wrong perception of social and economic realities [1]. Even students in statistics courses are reluctant to partake in them because they consider them complicated [5, 6]. Introductory statistics courses are an important part of various disciplines, but neither many teaching approaches have been noted nor the public understanding of

6th International GamiFIN Conference 2022 (GamiFIN 2022), April 26-29 2022, Finland
EMAIL: zampeta.legaki@tuni.fi (N.-Z..Legaki); daniel.fernandezgaleote@tuni.fi (D. Fernández.Galeote); juho.hamari@tuni.fi (J. Hamari)
ORCID: 0000-0002-2707-8364(N.-Z. Legaki); 0000-0002-5197-146X (D. Fernández Galeote); 0000-0002-6573-588X (J. Hamari)

these topics has been sufficiently advanced, making research in this direction valuable [3].

These gaps in pedagogical approaches related to engagement and motivation have recently spurred significant interest towards employing design principles from games as a pedagogical avenue. Games and gamification have been linked with intrinsic motivation, so they can be beneficial in an educational context [7]. The online and virtual educational environments that the COVID-19 pandemic forced to use make the integration of game-based learning in education even more appealing. Gamification, or enhancing a product or service by providing an experience like those afforded by games, has already been applied in statistics education with mostly positive results [8]. Nevertheless, literature reviews on this topic call for more empirical research and rigorous comparison among different gamified strategies to identify methods that effectively attract users' interest and support education. Empirical research on the effects of individual or simple motivational affordances and their comparison regarding learning outcomes [9, 10] is also crucial for understanding the effects of gamification and its integration in education.

Therefore, in this study, we investigate the effects of gamification on learning outcomes in the context of data literacy. We developed a web-based application that presents statistical concepts of Exploratory Data Analysis (hereafter *EDA*) using a variety of charts and real data. This application supports three additional versions, one for each of the following gamification types: challenge-, immersion-, social-based [11, 10]. We conducted a series of online experiments supporting random assignment of students (N=181) to one of the treatments (i.e., control, challenge-, immersion- and social-based gamification), using an online pre- and post-test experimental design. Students come from different courses, schools, and educational levels. This study aims to present an interactive educational application which teaches basic *EDA* topics, in a data literacy context, using real data sets about societal challenges, and to investigate the impact of different gamification types on learning outcomes.

## 2. Background

### 2.1. Data Literacy

Our society is becoming more data-rich every day. The penetration of digital technology in our daily life has accelerated the digitization and datafication of our society. These have shed light on the necessity of critical education about data as a set of strategies to support individuals in being aware, and understanding their data [12], because data is as useful as the insights, we can get from it [13]. Data literacy has been suggested as a solution leading to a data-driven society [14]. However, there is not yet a convergence regarding a definitive set of strategies to achieve data literate adults. Data literacy is described broadly as "*a set of abilities around the use of data as part of everyday thinking and reasoning for solving real-world problems*" [15] or "*the ability to understand and use data effectively (to inform decisions)*" [13, 16], and it is a crucial life skill nowadays [15], from finding employment to supporting decision-making [4]. However, data literacy is still in its infancy. Recent studies argue that it should be composed of data understanding and data use, but there is no consensus yet about its components. Other studies tie data literacy with statistical literacy [17] and claim the need for public understanding of statistics as a rising societal need to deal with a series of mis-es (e.g., misinformation, misunderstandings, etc.) [3, 18]. In this regard, statistical literacy, or "*the ability to interpret, critically evaluate, and communicate about statistical information*" [19], has been suggested as a desired goal for different educational levels and professions (e.g., researchers, various practitioners) too [3, 6]. Acknowledging that there is no universal definition of data literacy, for the needs of this study we use the broad definition referred to above, supported by the concept of statistical literacy, and see it as an evolving concept that helps individuals to use data as part of everyday thinking, get insights from a data-rich society by using common statistical tools, and supports solving real-world problems [14].

We are bombarded with statistical news and data daily, albeit statistics or related fields remain too complicated and difficult to engage with for adult learners [3, 20, 5]. Considering the range of disciplines that provide at least an introduction to statistics course, the importance of data literacy skills, and the need for data literate students and citizens [21], individuals' reluctance to engage with statistics is an urgent problem. In addition, the content of data or statistical literacy courses needs to be improved by including real data sets [22] and linking statistical concepts with everyday life topics [23, 24]. A few initiatives have been suggested to promote data or statistical literacy in

lifelong learning with promising results, e.g., the use of digital technology, workshops with real problems, and data and game-based/gamified activities that harness creativity [15, 25, 26, 2]. However, their limited number, along with a lack of research on data literacy pedagogy [12, 27], set the basis for exploring gamification's potential in this topic.

## 2.2. Gamification and data literacy

Hence, this study examines gamification's potential given its mainly positive impact in education. Gamification, defined as an "*intentional process of transforming any activity, system, service … into one which affords positive experiences, skills, and practices similar to those afforded by games*" [28], has been implemented in a variety of fields, and most empirical studies focus on education [10, 29]. Specifically, gamification has been mostly employed in Computer Science, Mathematics and Engineering [30, 31], noting mainly positive results regarding psychological and behavioral outcomes [10]. Gamification has gained acceptance in e-learning educational environments [32] as well, which makes it a useful addendum for virtual learning.

There are also studies in favor of gamification in the data and statistics fields, especially in digital formats [8, 33]. Most of these studies focus on introductory statistics courses, which include topics related to data and statistical literacy skills (e.g., basic *EDA* topics [40], chart interpretation). These courses are often taught in various disciplines. Thus, the need to motivate students becomes crucial. Other initiatives to increase student motivation in these fields also use interactive software or persuasive data visualization [35, 36], with mostly optimistic results. Despite that most of the studies list the positive impact of gamification or persuasive data visualization, their effects seem to vary depending on the context and the audience [37].

Gamification is sometimes categorized into three types based on the motivational affordances used, i.e., challenge/achievement (focusing on the feeling of competence and using points, badges, etc.), immersion (putting emphasis on avatars and narrative), and social (concentrating on competition and/or collaboration) [11, 10, 38]. [11] attempted to link these gamification types with different intrinsic needs satisfaction, i.e., autonomy, competence, and relatedness needs, finding differences among their effects but concluding to a positive impact of gamification, overall. Despite the variety of motivational affordances, still most of the studies focus on the triad of points-leaderboards-badges [10, 29, 37]. On top of that, negative results in education [39, 37] call for more research and cautious design of gamification in the education process. [31, 10, 9] also mention the need for more controlled empirical studies or studies that empirically compare individual or different types of motivational affordances and further examine the conditions in which gamification becomes effective in different contexts. Combining the above-mentioned need to evaluate gamification impacts and the need for increasing and improving teaching methods regarding data literacy, within this study we designed and implemented a system that supports three gamified versions (i.e., challenge-, immersion-, social-based gamification), along with a control version, that aims to teaches data literacy concepts through interactive charts and real data. Next, we conducted random assignment experiments using this system to address the above-mentioned gap.

# 3. Methods and data

## 3.1. Participants

We conducted online experiments in four different schools. Schools were recruited based on the first author's research network, and teachers' agreeing upon participant incentives with the first author. The total sample is composed of N=181 (58.56% male; 40.89% female; 0.05% non-binary or other) university students. The participants were students in different courses and schools, with different educational levels (undergraduate, postgraduate, and MBA students) as follows:

- 62 students; *Forecasting Techniques*; School of Electrical & Computer Engineering, National Technical University of Athens, Greece; class 2021; 4$^{th}$ year of undergraduate studies (FT ECE-NTUA).
- 18 students; *Experimental Research Methods*; Business Administration Department, University of Thessaly, Greece; class 2021; 2$^{nd}$ year of MBA studies (RM MBA-UTH).
- 36 students; *Quantitative Methods in Decision Making*, Business Administration Department, University of Thessaly, Greece; class 2021; 2$^{nd}$ year of MBA studies (DS MBA-UTH).
- 22 students; *Multimedia and Hyper-media Theory*; University of Pretoria, South Africa,

class 2021; 2nd year of undergraduate studies (MHT UPR).

- 30 students; *Forecasting and Data Analytics via Gamification*, Summer School, Tampere University, Finland; class 2021; under/postgraduate; (FDAG TAU).
- 13 students; *Special Issues in the Time Project Management*; Business Administration Department, University of Thessaly, Greece, class 2021, 4th year of undergraduate studies (PMUTH).

## 3.2. Materials

All materials were designed and implemented to meet the goals of this study, i.e., to effectively teach basic *EDA* concepts and compare the effects of different gamification types on learning outcomes. The materials of our experiments are composed of three main parts: a pre-test on-line questionnaire, a web-based educational application supporting four different versions (a control, and three versions of gamified learning), and a post-test online questionnaire. Every version of the web-based application matches to one of the following: control (i.e., no gamification elements), challenge- (i.e., badges), immersion- (i.e., avatars and a story related to the presented data), and social-based (i.e., illustrated text about the participant's rank among others) gamification. All the materials are in English. A more detailed description for each part follows.

### 3.2.1. Pre- and post-test

The pre- and post-tests have the same structure, number of questions, and topics. They are com-posed of 30 multiple choice questions related to *EDA* topics and misconceptions regarding worldwide data. Even though data literacy is gaining more importance and there are some courses available online, our search of standardized tests about data literary skills did not conclude to any result. Thus, a 30-question test was constructed by reviewing related online courses, data literacy and *EDA* literature [40, 34, 41], using part of Analytics Vidhya's test[2] (a few questions from a test about fundamental statistics skills), including data about the UN 2030 Sustainable Development Goals (hereafter SDGs)[3], and considering our expertise in statistics and forecasting. We concluded to the following structure regarding both *EDA* and data related topics: central tendency (3 questions), spread (5 questions), growth rate (2 questions), graph interpretation and data knowledge about SDGs (14 questions), re-expression of data and COVID-19 pandemic (2 questions), and regression and correlation (4 questions). Calculating the mean among some numbers or answering about the percentage of countries with laws against sexual harassment at work are some examples. We also included 2 attention questions. The order of the questions and answers and the description of some questions might slightly differ, but the questions related to the SDGs and the calculations needed to answer the pre- and post-tests are the same. The content of the questions is interwoven with the learning objectives and the content of the application. The pre- and post-tests are exactly the same for all the different versions.

### 3.2.2. Description of the (gamified) educational application

We designed the content and then implemented a publicly available web-based application from scratch, which aims to teach basic *EDA* [40, 34, 41] and compare different gamification types (i.e., challenge-, immersion-, social-based) regarding the learning outcomes. A brief description of the content and the design of the application follows:

**Content design**. The main *EDA* topics were chosen according to our learning objectives, well-known relevant literature [46, 34, 41], our target audience (i.e., adults), time limitations (a full round should last for approximately 1 hour), and the conditions of an online experiment. We opted for an online system to have students from various educational background and levels because of the COVID-19 pandemic restrictions. Our main goal was to provide data literacy skills to a broad audience, even to those without a strong statistical background. We borrowed some of the thematic axes from online courses, combining them with relevant literature [40, 34, 41] and our expertise. The web-based application is divided into 5 discrete pages/levels, and every page is linked with one topic as follows: page 1: central tendency; page 2: the spread of data; page 3: chart interpretation; page 4: re-expression of data; page 5: regression and correlation. The higher the level, the more complicated the topic is, integrating a

[2] https://www.analyticsvidhya.com/

[3] https://sdgs.un.org/goals

scaffolding approach [42]. For each level, an extra motivational affordance appears (see Figure 1). In addition, we included small interactive exercises as practice in four out of the five pages and colorful buttons at the top of every page that explained important topics using language as plain as possible. Figure 1 shows the 3rd page of the application, the colorful buttons with further explanation upon clicking on them, and an interactive chart with an in-application question. Moreover, we selected data related to the SDGs and other societal challenges such as the COVID-19 pandemic for the in-application examples and charts to increase engagement with these issues. Online open data sources were used, including *Our World in Data*, *The World bank*, *GAPMINDER*'s database and report, and others mentioned in the application.


**Figure 1:** Screenshot from the gamified application, immersion-based version.

**Gamification design**. One of the main goals of this study is to compare the three gamification types [10, 11] in an online application along with having no gamification, regarding *EDA* and data literacy learning outcomes. The distinction between gamification types is based mainly on different player motivational directions and the game elements used, and it has been associated with slight differences in need satisfaction [11]. Following the design guidelines by [43], we conducted a brainstorming session with six gamification experts from the authors' research group interested in gamification design for data literacy. Mainly, we focused on the integration of one motivational affordance per type to provide a game-like experience to participants with different stimuli but keep the same settings and interface whenever possible to have comparable versions. More iterations took place among the first and third authors of this study. We decided that the motivational affordance should be represented at the top and bottom of the page (see gamification placeholder, Figure 1), include verbal feedback, and updated, by adding one more, for each page. Regarding challenge-based gamification, we decided to include badges (i.e., bronze, silver, gold, and ruby stars). If the user's answers met the criteria for each page, then a new badge was added in the gamification placeholder. For the immersion-based version, a series of avatars related to the sustainable development agenda pillars (i.e., people, prosperity, planet, peace, and partnerships) were created, along with an illustrated story related to the SDGs. Figure 1 shows the illustrated story for a user who has completed page 3, so three stages of the story have been opened. The story progresses based on user correct answers in the application. Social-based gamification was presented via a competition element materialized through illustrated quotes and messages indicating the user's rank as compared to others. So, having completed the tasks for every page, the user discovered or lost an additional badge (challenge version), contributed or not to the sustainability goals based on the story (immersion version), and improved or not their ranking (social version). The spaces dedicated to the pictures and texts were blank for the control version.

**A full round in the application**. Initially, the user reads navigation instructions according to the version that they have been randomly assigned to. For the challenge-, immersion-, and social-based gamification there are additional descriptions regarding the respective elements. Additionally, for the immersion-based gamification, the user selects an avatar. Saving their choice, a user moves to page 1. Every page is composed of 4 to 10 interactive charts, a question for each, and interactive calculators to help answering the questions. For each page, participants should correctly answer an increasing number of questions to meet the criteria and gain the extra respective badge, or contribute to the story that they participate in, or upgrade their rank. The cumulative number of participants' correct answers defines the competitive messages they receive regarding their rank. A participant needs to answer all the questions for each page to proceed to the next one, and all the completed pages remain available. Feedback regarding

correct answers is available only for the completed levels, with the respective motivational affordance per gamification type. A full round is done when the tasks in page 5 are completed. Then, the user reaches the post-test. During the round, users can logout (progress is saved). A user who has completed a full round cannot participate again.

### 3.2.3. Procedure and experimental design

The experiments were conducted in the context of six courses at different schools. All the participants received the same instructions regarding the application. However, students in FT ECE-NTUA, DS MBA-UTH, RM MBA-UTH, and PM UTH had the instructions in Greek and the incentive for participation was a bonus of 1 out of 10 in the course's final grade instead of an equivalent exercise in the final exam. Students in FDAG TAU and MHT UPR received the instructions in English and their participation was mandatory as part of the course. Participants were instructed to use a computer. They were aware of the possibility to logout and sign in and the time available for participation varied per school. Students in FT ECE-NTUA had a month available to register and complete a full round, students in MHT UPR, and DS MBA-UTH, RM MBA-UTH, and PMUTH had two weeks, and students in FDAG TAU had one week. These differences are due to the different course settings.

All participants had to register and give informed consent to proceed. Upon their registration, they had to complete all the pre-test questions, which were not available afterward. There was no feedback regarding the pre-test questions. Having saved their answers, they were randomly assigned to one of the four conditions, i.e., control, challenge-, immersion-, or social-based gamification, using the *sample()* function of the R-base package. Then, participants had to read the instructions. Additional instructions were provided to participants based on the version that they had been assigned to. For example, assigned participants to the immersion version had to read additional instructions regarding the game elements and select one of the avatars.

Having read the instructions and independently of the version, participants were directed to page 1, where they needed to answer questions based on the provided charts and/or use calculators. Then, they got feedback about their choices. Participants assigned to one of the gamified conditions had one new gamification element (a badge, a strip of a story, or their illustrated rank) available at the top and bottom of every page, based on their performance. The same process was followed for all the available pages, having different datasets/charts and tasks per page, up to page 5. All the previous levels along with correct answers were available while participants moved to the next pages. Having saved their answers on page 5, participants were directed to the post-test. All participants had to answer the same post-test questions. All the previous pages (apart from the pre-test) were available without the correct answers. Figure 2 illustrates the experimental design and the procedure that was followed.

## 4. Results

The objective of this study is to present and evaluate an online application which uses real worldwide data to teach basic data literacy topics and investigate the impact of three gamification types regarding the learning outcomes. Hence, we collected students' performance in pre- and post-tests. Student performance for each test was calculated as the sum of the correct answers. Considering all the questions as equivalent, the maximum score per test is equal to the number of questions, i.e., 30. The overall statistical approach to this study's results is divided into two steps. Initially, the collected data from both pre- and post-test questionnaires are examined using descriptive statistics to explore the group means, standard deviations, and numbers. Figure 3 illustrates the performances of students per test and treatment and Table 1 presents the descriptive statistics per group and school, too.

In terms of an overall evaluation of the educational application, the mean value of students' post-test performance (M=19.03, SD=5.05) is higher than their pre-test performance (M=13.24, SD=4.09), as expected. We conducted a paired t-test, with a confidence interval equal to 95%. The null hypothesis: $H_0$ *equal differences in means* is rejected ($t = 20.634$, $df=180$, $p<0.001$), thus using the suggested educational application improves mean performance in the context of data literacy by 43.73%, resulting in a large effect size ($d=1.26$).

**Table 1**
Students' performance per different gamification types (pre- vs post-test).

| Group | N | Pre-test | | Post-test | | Difference | | Improvement | Wilcox sign test | Effect size |
|---|---|---|---|---|---|---|---|---|---|---|
| | | M | SD | M | SD | M | SD | | | |
| Control | 48 | 13.29 | 3.57 | 19.15 | 4.78 | 5.85 | 3.86 | 44.04% | Z=19.5 p<0.001 | r=0.839 (large) |
| Challenge- | 46 | 12.72 | 4.23 | 17.67 | 5.38 | 4.96 | 3.83 | 38.97% | Z=29.5 p<0.001 | r=0.817 (large) |
| Immersion- | 55 | 13.58 | 4.52 | 19.13 | 5.16 | 5.55 | 3.83 | 40.83% | Z=13.5 p<0.001 | r=0.842 (large) |
| Social- | 32 | 13.34 | 3.95 | 20.62 | 4.44 | 7.28 | 3.1 | 54.57% | Z=0 p<0.001 | r=0.875 (large) |

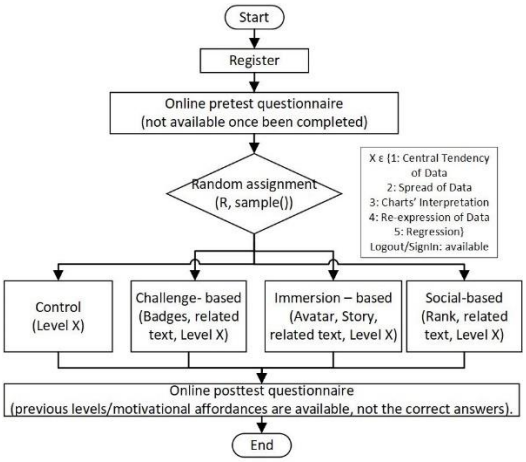


**Figure 2:** Flowchart of the experimental design.

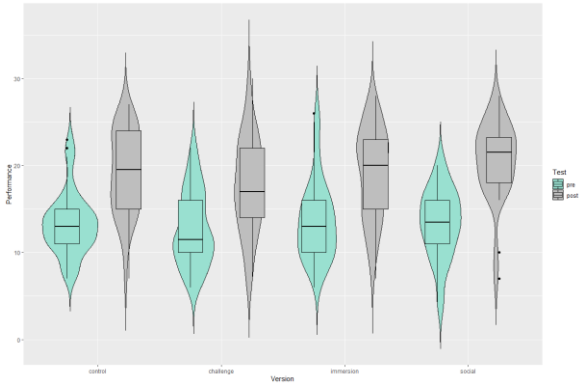


**Figure 3:** Student performance in pre- and post-tests per gamification type.

In order to further examine the differences between pre- and post-tests in the different gamification strategies, we opted for non-parametric tests, due to the violation of the normal distribution assumption in the groups. A Wilcoxon Signed-Ranks Test for each type indicates that the mean post-test ranks were statistically significantly higher than the mean pre-test rank for each type of gamification, with a confidence interval equal to 95%. Table 1 presents mean values, standard deviations, numbers for each type, differences in pre- and post-test performance and the improvement along with respective Wilcoxon effect sizes.

Next, the impact of different gamification types regarding learning outcomes in *EDA* topics is examined. Analysis of covariance (ANCOVA) was chosen to examine the effects of using different gamification types on student performance, controlling for initial differences in the pre-test. However, having different group sizes and not meeting the assumption of linear relationship between the dependent variable and the covariate, we follow the non-parametric alternative, using the *sm* and *fANCOVA* packages in *R (v3.5.3)* for validation. Four curves have been calculated based on polynomial regression with automatic smoothing parameter selection via AICC for curve fitting. Based on the comparison of four non-parametric regression curves, the null hypothesis "$H_0$*: there is no difference between the 4 curves*" cannot be rejected (T=21.08, p=0.741).

Acknowledging non-parametric analysis limitations, we also conduct a one-way ANOVA on performance change. Our sample meets the ANOVA assumption (i.e., normal distribution, homogeneity of variance, and the observations are independent of each other). No statistically significant differences were detected F(3,177)= 2.563, p=0.056. This fact is in line with the non-parametric analysis, and it implies that all groups showed a similar learning gain for all the different gamification types, including the control group, regarding student performance change.

## 5. Discussion and conclusions

### 5.1. Discussion

Overall, the results suggest that the use of the online application improved learning outcomes regarding data literacy, as investigated in this study, using interactive charts and tools, with real data sets related to current societal challenges, i.e., COVID-19 and SDGs. Other studies suggest that learning objectives in the context of data literacy can be achieved by using data visualization techniques and statistics as persuasive technology

means, that is as interactive technology that aims to change a person's attitudes or behavior [44]. They can also promote critical thinking among students [45,46] and provide opportunities in teaching-learning process [47,48]. Despite that the use of persuasive technology, or captology as is mentioned [44], is not yet mature in education, there are some preliminary positive indicators about its potential on some learning variables such as attitude and motivation, but further research is needed [49]. Hence, the noted improvement of all the versions, control included, could be an effect of the interactive charts and tools provided and the chosen thematic areas, which could impact on motivation positively and eventually learning. This finding could be in accordance with [35], who examine the effects of data visualization as a persuasive visualization tool that might positively impact people's attitude and memorization or even improve accuracy as in a Bayesian reasoning problem [36]. However, further empirical research is needed in this area [35].

Another important finding is that the integration of different gamification types did not result in statistically significant differences on students' learning outcomes. Despite the mentioned positive effects of gamification in education [10, 8], there are a few studies commenting on the potential negative effects of gamification [37]. Based on [37], 35% of the reviewed papers mentioned indifference as an effect, when gamification did not impact for better nor for worse. Our results are in line with [50, 51], where there was no significant impact of gamification on e-learning interventions on students' performance, even though in these studies the participants' initial motivation was higher and the described interventions lasted longer. In our study, the novelty effect or research fatigue might contribute also to this lack of effect on performance since most of the students completed the full activity, on average, in two hours, rather than logging out and signing in.

## 5.2. Limitations

There are some limitations regarding the design of the application. All the versions, control included, contain interactive charts, icons/emojis, and colorful buttons. The control version does not include any gamification, but it might be playfully framed given its interactivity, diverse colors, and a user-friendly design. So, even the control version might afford a playful experience. Even though badges, avatars and a story, and competition are representative of the challenge-, immersion-, and social-based gamification [10], our results are limited regarding the implementation, the sample, and the described context.

Another limitation refers to the sample and the procedure of the experiments. The sample sizes, the difference in students' schools, and years of study within the students' distribution into different conditions might affect the homogeneity of slopes between pre- and post-test' s performances. Since a non-parametric analysis was conducted, the validity of the results is not affected. The difference in incentives needs to be mentioned, as well. Students in FT ECE-NTUA, RM MBA-UTH, DSMBA-UTH, PM UTH received 1 point out of 10 as a bonus in their grade, instead of an equivalent exercise at the end. However, students in FDAG TAU and MHT UPR participated in the application as part of mandatory assignments to successfully pass the course. Finally, despite the difference in the instructions about the available time to complete a full round, all students but one completed a full round in a maximum of three days. This study focuses only on the impact of a web-based application on data literacy and the comparison among different gamification types. However, both pre- and post-test questionnaires comprise more questions than the knowledge questions, which might lead to research fatigue. A larger sample is suggested, and completing a full round during three days, even though the noted improvement shows the potential of this approach.

## 5.3. Conclusions

In our study, a (gamified) application was designed and implemented to teach data literacy and compare the impact of different gamification types on learning outcomes. Our results indicate an average of 43.73% improvement in learning outcomes and suggest optimism regarding the contribution of interactive data visualization, interactive tools, and a friendly user interface in improving data literacy. However, we should be more skeptical about the integration of gamification when there is already a system with these characteristics as a basis. Employing a larger sample of the general public will strengthen the results and support data literacy teaching. In addition, investigating gameful experience constructs and connecting the used gamification

features with the improvement in specific learning outcomes and data literacy topics will provide insightful perspectives regarding the impact of gamification design choices in data literacy.

## 6. Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie, grant agreement No 840809, the Academy of Finland Flagship Program (337653 Forest-Human-Machine Interplay (UNITE)) and the Nessling Foundation (project No 202100217).

## 7. References

[1] H. Rosling, Factfulness, Flammarion, 2019.
[2] C. D'Ignazio, R. Bhargava, Databasic: Design principles, tools and activities for data literacy learners, The Journal of Community Informatics 12 (2016).
[3] F. C. Von Roten, Do we need a public understanding of statistics?, Public Understanding of Science 15(2006) 243–249.
[4] A. Yoon, A. Copeland, P. J. McNally, Empowering communities with data: Role of data intermediaries for communities' data utilization, Proceedings of the Association for Information Science and Technology 55 (2018) 583–592.
[5] M. D. Albritton, P. R. McMullen, Classroom integration of statistics and management science via forecasting, Decision Sciences Journal of Innovative Education 4 (2006) 331.
[6] D. J. Rumsey, Statistical literacy as a goal for introductory statistics courses, Journal of statistics education 10 (2002).
[7] N.-Z. Legaki, K. Karpouzis, V. Assimakopoulos, J. Hamari, Gamification to avoid cognitive biases: An experiment of gamifying a forecasting course, Technological Forecasting and Social Change 167 (2021) 120725.
[8] Z. Legaki, J. Hamari, Gamification in statistics education: A literature review (2020).
[9] L. E. Nacke, C. S. Deterding, The maturing of gamification research, Computers in Human Behaviour (2017) 450–454.
[10] J. Koivisto, J. Hamari, The rise of motivational information systems: A review of gamification research, International Journal of Information Management 45 (2019) 191–210.
[11] N. Xi, J. Hamari, Does gamification satisfy needs? a study on the relationship between gamification features and intrinsic need satisfaction, International Journal of Information Management 46 (2019) 210–221.
[12] L. Pangrazio, J. Sefton-Green, The social utility of 'data literacy', Learning, Media and Technology 45 (2020) 208–220.
[13] M. Frank, J. Walker, J. Attard, A. Tygel, Data literacy-what is it and how can we make it happen?, The Journal of Community Informatics 12 (2016).
[14] T. Koltay, Data literacy: in search of a name and identity, Journal of Documentation (2015).
[15] A. Wolff, D. Gooch, J. J. C. Montaner, U. Rashid,G. Kortuem, Creating an understanding of data literacy for a data-driven society, The Journal of Community Informatics 12 (2016).
[16] E. B. Mandinach, E. S. Gummer, A systemic view of implementing data literacy in educator preparation, Educational Researcher 42 (2013) 30–37.
[17] M. Shields, Information literacy, statistical literacy, data literacy, IASSIST quarterly 28 (2005) 6–6.
[18] K. K. Wallman, Enhancing statistical literacy: Enriching our society, Journal of the American Statistical Association 88 (1993) 1-8.
[19] I. Gal, Adults' statistical literacy: Meanings, com-ponents, responsibilities, International statistical review 70 (2002) 1–25.
[20] J. B. Ramsey, Why do students find statistics so difficult, Proceedings of the 52th Session of the ISI. Helsinki (1999) 10–18.
[21] F. C. von Roten, Y. de Roten, Statistics in science and in society: From a state-of-the-art to a new research agenda, Public Understanding of Science22 (2013) 768-784.
[22] R. W. Erwin Jr, Data literacy: Real-world learning through problem-solving with data sets, American Secondary Education (2015) 18-26.
[23] B. Berikan, S. Özdemir, Investigating "problem-solving with datasets" as an implementation of computational thinking: A literature review, Journal of Educational Computing Research 58 (2020) 502–534.
[24] A. Yoon, A. Copeland, Understanding social im-pact of data on local communities, Aslib Journal of Information Management (2019).

[25] S. Werning, Making data playable: A game co-creation method to promote creative data literacy., Journal of Media Literacy Education 12 (2020) 88–101.
[26] R. Bhargava, C. D'Ignazio, Designing tools and activities for data literacy learners, in: Workshop on Data Literacy, Webscience, 2015.
[27] A. Wolff, M. Wermelinger, M. Petre, Exploring design principles for data literacy activities to sup-port children's inquiries from complex data, Inter-national Journal of Human-Computer Studies 129(2019) 41–54.
[28] J. G. Hamari, G. Ritzer, C. Rojek, The Blackwell encyclopedia of sociology, 2019.
[29] M. Trinidad, M. Ruiz, A. Calderón, A bibliometric analysis of gamification research, IEEE Access 9(2021) 46505-46544.
[30] J. Swacha, State of research on gamification in education: A bibliometric survey, Education Sciences11 (2021) 69.
[31] C. Dichev, D. Dicheva, Gamifying education: what is known, what is believed and what remains uncertain: a critical review, International journal of educational technology in higher education 14 (2017)1-36.
[32] A. N. Saleem, N. M. Noori, F. Ozdamli, Gamification applications in e-learning: A literature review, Technology, Knowledge and Learning (2021) 1–21.
[33] A. Lekka, E. Toki, C. Tsolakidis, J. Pange, Literature review on educational games for learning statistics, in: 2017 IEEE Global Engineering Education Conference (EDUCON), IEEE, 2017, pp. 844–847.
[34] J. W. Tukey, et al., Exploratory data analysis, volume 2, Reading, Mass., 1977.
[35] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, E. Bertini, The persuasive power of data visualization, IEEE transactions on visualization and computer graphics 20 (2014) 2211–2220.
[36] L. Micallef, P. Dragicevic, J.-D. Fekete, Assessing the effect of visualizations on bayesian reasoning through crowdsourcing, IEEE transactions on visualization and computer graphics 18 (2012)2536–2545.
[37] A. M. Toda, P. H. Valle, S. Isotani, The dark side of gamification: An overview of negative effects of gamification in education, in: Researcher links workshop: higher education for all, Springer, 2017, pp. 143–156.
[38] N.-Z. Legaki, N. Xi, J. Hamari, K. Karpouzis, V. Assimakopoulos, The effect of challenge-based gamification on learning: An experiment in the context of statistics education, International journal of human-computer studies 144 (2020) 102496.
[39] M. D. Hanus, J. Fox, Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance, Computers& education 80 (2015) 152–161.
[40] M. Komorowski, D. C. Marshall, J. D. Salciccioli, Y. Crutain, Exploratory data analysis, Secondary analysis of electronic health records (2016) 185–203.
[41] R. K. Pearson, Exploratory data analysis using R, CRC Press, 2018.
[42] M. Stewart, Understanding learning: theories and critique, in: University teaching in focus, Routledge, 2012, pp. 3–20.
[43] B. Morschheuser, L. Hassan, K. Werder, J. Hamari, How to design gamification? a method for engineering gamified software, Information and Software Technology 95 (2018) 219–237.
[44] B. J. Fogg, Persuasive computers: perspectives and research directions, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 1998, pp. 225–232.
[45] S. Forbes, J. Chapman, J. Harraway, D. Stirling, C. Wild, Use of data visualisation in the teaching of statistics: A new zealand perspective., Statistics Education Research Journal 13 (2014).
[46] L. Ryan, Visualization techniques to cultivate data literacy, in: Advances in exemplary instruction, CreateSpace, 2015.
[47] S. Devincenzi, V. Kwecko, F. P. de Toledo, F. P. Mota, J. Casarin, S. S. da Costa Botelho, Persuasive technology: Applications in education, in: 2017 IEEE Frontiers in Education Conference (FIE), IEEE, 2017, pp. 1-7.
[48] B. J. Fogg, Persuasive technology: using computers to change what we think and do, Ubiquity 2002(2002) 2.
[49] S. Agnisarman, K. C. Madathil, L. Stanley, A survey of empirical studies on persuasive technologies to promote sustainable living, Sustainable Computing: Informatics and Systems 19 (2018) 112–122.
[50] P. M. Papadopoulos, T. Lagkas, S. N. Demetriadis, How revealing rankings affects

student attitude and performance in a peer review learning environment, in: International Conference on Computer Supported Education, Springer, 2015, pp. 225-240.
[51] A. Domínguez, J. Saenz-de Navarrete, L. De-Marcos,L. Fernández-Sanz, C. Pagés, J.-J. Martínez-Herráiz, Gamifying learning experiences: Practical implications and outcomes, Computers & education 63(2013) 380–392.